The U.S. water data gap: A survey of state-level water data platforms to inform the development of a national water portal

Article:

Josset, Laureline, Allaire, Maura, Hayek, Carolyn, Rising, James, Thomas, Chacko and Lall, Upmanu (2019) The U.S. water data gap: A survey of state-level water data platforms to inform the development of a national water portal. Earth's Future, 7 (4). pp. 433-449. ISSN 2328-4277

https://doi.org/10.1029/2018EF001063

Earth's Future

RESEARCH ARTICLE
10.1029/2018EF001063

The U.S. Water Data Gap—A Survey of State-Level Water Data Platforms to Inform the Development of a National Water Portal

Laureline Josset[1], Maura Allaire[2], Carolyn Hayek[1], James Rising[3], Chacko Thomas[4], and Upmanu Lall[1]

[1]Columbia Water Center, Columbia University, New York, NY, USA, [2]Department of Urban Planning and Public Policy, University of California, Irvine, CA, USA, [3]Grantham Research Institute, London School of Economics, London, UK, [4]School of Regulation and Global Governance, Australian National University, Canberra, ACT, Australia

Abstract Water data play a crucial role in the development and assessment of sustainable water management strategies. Water resource assessments are needed for the planning, management, and the evaluation of current practices. They require environmental, climatic, hydrologic, hydrogeologic, industrial, agricultural, energy, and socioeconomic data to assess and accurately project the supply of and demand for water services. Given this context, we provide a review of the current state of publicly available water data in the United States. While considerable progress has been made in data science and model development in recent years, data limitations continue to hamper analytics. A brief overview of the water data sets available at the federal level is used to highlight the gaps in readily accessible water data in the United States. Then, we present a systematic review of 275 websites that provide water information collected at the state level. Data platforms are evaluated based on content (ground and surface water, water quality, and water use information) along with the analytical and exploratory tools that are offered. Wev discuss the degree to which existing state-level data sets could enrich the data available from federal sources and review some recent technological developments and initiatives that may modernize water data. We argue that a national water data portal, more comprehensive than the U.S. Energy Information Administration, addressing the significant gaps and centralizing water data is critical. It would serve to quantify the risks emerging from growing water stress and aging infrastructure and to better inform water management and investment decisions.

Plain Language Summary Water data are essential to describe the state of our resources. They enable the assessment of risks, the evaluation of management decisions, and the design of infrastructure. To adequately manage water, it is necessary to have information not only on climate, the environment, the rivers, and the aquifers but also on the dams and canals that store and divert water, the use of water, and the laws that affect the attribution and distribution of water. Current data limitations stand out as the Achilles heel to promote a sustainable and resilient water management. After a brief overview of the shortcomings of the current readily accessible federal water data, we present results from a systematic review of 275 websites hosting water information collected at the state level. We categorize and evaluate the data platforms based on content (ground and surface water, water quality, water use, and water finance) along with the tools that are offered to interpret and make use of the data. We conclude with a discussion on the degree to which the reviewed data sets at the state level could address the gap in federal data sets. We argue that a national water data portal is critical to understand the growing water risks from extensive use, climate, and aging infrastructure and to better inform water management.

Key Points:
- We discuss key data gaps in ground and surface water, quality, use, and financial data that hinder water management in the United States
- We review state-level approaches to online platform for water data, analytics, visualization, and projections, identifying content and tools
- We conclude with recommendations for the development of a national water data portal to inform public and private decisions

Supporting Information:
- Supporting Information S1
- Data Set S1

Correspondence to:
L. Josset,
lj2390@columbia.edu

Citation:
Josset, L., Allaire, M., Hayek, C., Rising, J., Thomas, C., & Lall, U. (2019). The U.S. water data gap—A survey of state-level water data platforms to inform the development of a national water portal. *Earth's Future*. *7*, 433–449. https://doi.org/10.1029/2018EF001063

Received 2 OCT 2018
Accepted 18 FEB 2019
Accepted article online 1 MAR 2019
Published online 25 APR 2019

1. Introduction

Water data are the foundation of water research. While numerous advances have been made in the understanding of water processes, models and techniques are limited by the data available as inputs, whether it is in hydrology (surface water data), hydrogeology (groundwater), or hydroeconomics (use or financial data). Most water research, regardless of scale, mentions lack of data as a key limitation. Statistical techniques have been developed to deal with sparse and incomplete information. However, propagation of the uncertainty and error measurements through the models is often limited by this lack of data. The risk of biased conclusions is therefore not uncommon, and these issues are only exacerbated when attempting projections and forecasts.

Water data challenges extend beyond academia. Reliable data are central to the design and evaluation of public sector decisions (Gallaher & Heikkila, 1990; Laituri & Sternlieb, 2006). This is especially the case for the development and implementation of sustainable water management strategies (High Level Panel on Water, 2017). To accurately describe and predict both supply and demand requires the following: hydrological and hydrogeological data to understand the physical system; climatic data to quantify weather impacts on availability and use; and industrial, agricultural, energetic, and socioeconomic data to evaluate demands and responses. Thus, the challenge extends beyond the question of water allocation to questions of finances, infrastructure, and health associated with water security and climate variability, such as size of dams and reservoirs, design of water treatment plants, and set water rates. The World Economic Forum has cited water risk as a critical concern for businesses. Hence, the corporate sector is similarly dependent on water data to assess production, demand, and competition potential; establish resilient supply chains; and evaluate climatic risks and investments in water and other infrastructure. Water information is also useful for individuals, whether it is used to adapt their direct use of water, make informed decisions regarding the water footprint of their usage patterns, or assess water quality concerns and the associated health risk.

1.1. U.S. Water Data at the Federal Level

Renewed attention has been placed on the state of water data in the United States, as information is key to “fix water” (Fishman, 2016; Jerome, 2006). In the absence of data and rigorous analysis to establish the current state of the system, we do not know exactly what these issues are. The state of water use data in the United States has been referred to as primitive (Fishman, 2016; Jerome, 2006). Comprehensive nationwide water use estimates (Dieter et al., 2017; Maupin et al., 2017) are only produced every 5 years with their publication typically delayed; 2015 estimates for the public and domestic sectors were released in November 2017 (Dieter & Maupin, 2011) and for other sectors in June 2018 (Dieter et al., 2017), meaning that until then only 2010 estimates (Maupin et al., 2017) were available on which to base any sector-specific analysis. Estimates are produced at the county scale, a rather coarse resolution in particular in the more arid western states, where counties have large land area, and as annual averages, making it impossible to quantify crucial seasonal cycles.

Besides issues regarding the temporal and spatial scales of the measurements and the lag in their availability, water use estimates are prone to uncertainty, not only because of the difficulty in defining them (Ruddell, 2015) but also from measurement and reporting errors. In a recent analysis of the water data for thermoelectric plants, Harris and Diehl (2019) highlight discrepancies between U.S. Energy Information Administration data, U.S. Geological Survey (USGS) data, and USGS models on the order of 23% for the United States, though that sector's water use is arguably the easiest to estimate with few users and strong regulation. This cast doubt on our capacity to understand water use patterns and to forecast future water demand (Perrone et al., 2017), despite a more pronounced need as water supply fluctuations and periods of drought increase.

Data concerns are not restricted to water use. In 2001, Vörösmarty et al., 2014 described the whole landscape of water data as “a new endangered species” due to reduced funding and a decreasing number of gauges. With respect to groundwater information, the USGS has proposed new platforms and services (Blodgett et al., 2002; Hirsch & Fisher, 2017), but the uneven and limited spatial coverage of gauges (1,559 real-time ground water table gauges in 2014; Hirsch & Fisher, 2017) only provides a coarse overview. The spatial issue is not only on the latitude-longitude scale: Russo and Lall (2018) note that a lack of deep aquifer data limits a robust assessment of signals in water table depth as a function of climate or water use.

The availability of water quality information is equally troubling. A significant amount of data on chemical and biological attributes is collected from surface and ground water bodies, waste water streams, and drinking water suppliers. For raw water, the USGS operates a National Water Quality Assessment program that surveys benchmark sites for assessment (National Research Council, 2002). However, the capacity to assess trends in the national performance of regulations related to the Clean Water Act and the Drinking Water Act is rather limited. This is in part due to the limited sampling of the attributes and largely due to the quality of the relevant databases maintained by the federal agencies. The first national assessment of trends in drinking water quality (Allaire et al., 2018) was only presented in 2018, using the publicly available data sets from the U.S. Environmental Protection Agency (EPA) Safe Drinking Water Information System and separately available data sets from the U.S. Census Bureau. A major caveat for national drinking water quality information is that violations of Safe Drinking Water Act standards are underreported at the national level. The EPA and Government Accountability Office estimate that 26–38% of health-related violations are either not reported or inaccurately reported to the national Safe Drinking Water Information System (EPA, 2017; GAO, 2016). Similarly, regarding information related to Clean Water Act compliance, Sprague et al. (2017) report that nearly 58% of the records pertaining to nutrients in surface waters, as reported in the Water Quality Portal (WQP) and in records obtained directly from major water-resource agencies in each state, presented ambiguous or flawed metadata. These problematic records that represent data collection costs of $6.8 to $19 billion potentially prevent a secondary use of the data that are, in effect, wasted. Beyond ambiguous metadata, our study found a large fraction of the nutrient data in the WQP were inconsistent at the sample level. Our analyses of dissolved oxygen data for surface waters across the United States, as available in the WQP, also indicate that 28% of records contain either negative values or values that are physically unfeasible, even when accounting for unit conversion uncertainties. This lack of data for water quality was brought forward twice by the UN rapporteur for extreme poverty, in particular for Alabama (De Albuquerque, 2014; Walton, 2001), illustrating how dire the situation is.

The financial dimension of water issues is also of concern. Indeed, improper water quality is often ascribed to a lack of funds to address the issue. A national overview of the financial characteristics of water utilities is only available through three Community Water Systems surveys that were conducted in 1995, 2000, and 2006.

Individual utilities often use municipal bonds or similar vehicles to generate finances needed for major upgrades, expansion of service, and infrastructure renewal. Accordingly, there is limited ability to benchmark performance, raising the costs associated with obtaining private funds. Another challenge associated with inadequate benchmarking information is projections of future funding needs, which might be particularly important as delays in investments, lack of maintenance, and aging may translate in the simultaneous expiration of much of the water infrastructure. The Congressional Budget Office in its assessments of the need for a federal plan for water infrastructure in 2002 noted that due to the lack of representative data from utilities and future economic, regulatory, and technological conditions, their estimates could be outside of the estimated range of uncertainty (Beider & Tawil, Office2002). Highlighted by the recent Oroville dam issues, lack of data on reservoir levels, use, governance, and their maintenance are equally concerning, and though frightening on their own, they are only to be exacerbated in a context of climate change (Ho et al., 2017).

The lack of publicly available water data also impacts Environmental Justice issues. Specific concerns include the quality of water services, especially in the rural and poorer areas, and issues of affordability. The lack of a national database containing information on water prices and assistance programs at the community level prevents the systematic quantification of affordability and forces researchers to perform the aggregation of several data sets themselves or having to pay to access partial surveys conducted by the industry (Mack & Wrase, 2013). With limited or no data on these problems, they become not only unaddressable but also invisible (Fishman, 2016).

In addition to the lack of availability or poor frequency of nationwide data, the information is scattered across several institutions. While vast amounts of freshwater resources data are collected and archived by local, state, and federal departments and agencies, there is little coordination on how these data are collected, organized, or stored, leading to discrepancies in estimates (Averyt et al., 2013; National Research Council, 2012; Shaffer, 2017). Accordingly, the collection of the information necessary for the construction of a national water account required, until recently, the consultation of many websites, as acknowledged by the USGS (National Academy of Sciences, 2016). The National Water Information System (NWIS) consolidates data on quantity and quality of surface and ground water in one-spot web interfaces. Though this a key advance, it provides limited practicality for aggregated or large-scale analyses and no information on water use. In addition, the platform only allows for the selection and download of the data without offering further help as to how to interpret it. To avoid the pitfall of considering data a replacement for understanding and theoretical concepts (Crutchfield, 2016), the simple collection of data is not enough; data sets need to be usable and transparent.

The limitations and advancements of water data federal platforms are very much dependent on the data collected by state-level authorities. Indeed, state-level agencies are responsible for the monitoring and enforcement of national regulations (e.g., for the Clean Water Act and Safe Drinking Water Act), as well as state-level regulations (e.g., in-stream flows, water rights, allocations, and withdrawals limits). The role of the federal agencies, whether regulatory or not, are in great part limited to gathering and presenting state-level data sets (National Academy of Sciences, 2016). Motivated by the federal agencies' dependence on states to construct their data sets, we conducted a systematic evaluation of water data sets at the state level to assess the possibility of filling the federal water data gaps as well as to provide an overview of current practices in data centralization and dissemination efforts.

The remainder of the present paper is organized as follows. We first detail the methodology followed to survey data websites at the state level (section 2), before proceeding to the description and analysis of the collected results (section 3). We conclude with the potential of addressing the national water data gaps using state-level data and present our suggestions for a water data platform at the federal level to inform water management, investment, and development for the United States.

2. State-Level Data Survey Methodology

A systematic survey of state-level water information was conducted to identify (1) data sets that are present at the state level that could be aggregated to augment water information at the national scale and (2) examples that could be followed to offer better data visualizations and construct a centralized national water data platform. Scientific, public or private analyses of the water sector require many types of information. In order to present a more global overview without preferring one objective or one sector, we review any data sets that can inform the environmental or anthropic dimension of water.

2.1. Website Identification and Selection

The first step of the survey consists of the identification of websites hosting water data at the state level. This is performed by systematically exploring the 100 first results in web search engines for "[*state name*] water data" for the 48 conterminous United States. Variations were also attempted to identify specific categories of data when no results were found among the 100 first results.

The identified websites are then explored to extract the specific pages hosting the data sets. Information disclosed within reports are ignored as the format suggests that they contain aggregated or estimated numbers, as opposed to primary data. Furthermore, the format prevents direct integration with other models or applications. Note that state websites directly associated with a local USGS office are not retained as we attribute them to national-level information.

2.2. Web Page Evaluation

The selected web pages are then evaluated on data content, analytics presented, and available exploration tools, to assess both the raw data and the functionality offered. The process consisted of three steps:

(1) *Hosting institution*

The identity of the hosting institution is reported and categorized as governmental, academic, civil society, corporate, consortium, or media organizations.

(2) *Content determination*

A rating of the websites content is performed by identifying which type of data are disclosed, regardless of the completeness, coverage, and frequency of updates of the information. The five categories and their respective subcategories are the following:

1. *Surface water*—combining estimates of discharge and volume for rivers, lakes, and reservoirs and information consigned in surface water permits. Subcategories: gauge location, discharge, reservoir level, permit, and use type.
2. *Groundwater*—focusing on groundwater levels for both observation and use whether recorded through time or consigned in groundwater permit information or well records. Subcategories: well location and depth, water table depth, well record, permit information, and sector of use.
3. *Water quality*—referring equally to chemical and biological in-stream quality data and to post-treatment drinking quality measurements. Subcategories: groundwater quality measurements, surface water quality measurements, and return permits.
4. *Water use*—comprising withdrawals, consumption, and return flows of water. Subcategories: withdrawal volumes, diversion volumes, point of use, sector of use, and return volumes.
5. *Financial* data—considering any expense or price data sets pertaining to water, such as capital or variable costs for extraction, treatment, distribution, and waste collection, or purchase and contract prices.

(3) *Evaluation of the disclosure of the data*

Table 1
Number of States with Information for Each of the Five Categories of Data

Groundwater: 44	Surface water: 42	Quality: 38	Use: 22	Financial: 11
Wells location: 43	Gauges location: 37	Groundwater: 31	Withdrawals: 14	Rates: 10
Wells depth: 37	Discharge: 20	Surface water: 34	Diversion: 10	Infrastructure: 2
Water table: 35	Reservoirs level: 23	Return permits: 16	Point of use: 5	
Record or permit: 37	Permit: 24		Use type: 14	
Type of use: 33	Type of use: 20		Water use: 11	
			Return: flows 3	

Note. Number of states in conterminous United States presenting readily available information for each of the five categories and their respective subcategories of data.

For each category of content listed above, we first assess if analytics are proposed to facilitate the understanding and manipulability of the data, in particular *summary statistics*, *quantification of uncertainty*, *projections*, and *export capabilities* for further analysis. Additionally, we review how the data are disclosed. We remark in particular if *visualization* options are included (graphs and plots) and specifically if *maps*, interactive or not, are offered. The selected websites are rated according to these categories, here again without assessing the quality, relevance, and thoroughness of the tools offered but only by identifying the presence of these attributes.

2.3. Disclaimer

The first limitation of our survey is that we focus only on websites with readily available information. While undoubtedly much more information is collected which could be disclosed and used, we limited the scope of the present survey, to avoid contacting the responsible agencies or filing a Freedom of Information Act request, which constitutes an additional barrier to overcome for a water data user that is sometimes pursued by researches and met with mixed results.

Second, the survey task is difficult due to the great variability from one state to another in definition, content, format, vocabulary, and approach to data collection and management. The authors performed the selection and classification of the websites to the best of their ability but acknowledge that the process is prone to inconsistencies due to inherent variations from one platform to the other.

Last, the identification and evaluation of the websites were performed over the time period from July to October 2017. In addition to the disclosure of the survey results in the supporting information of this paper, we have created an online forum to maintain the survey table, keep it up to date, and augment its content when websites evolve or if new ones appear.

3. Results

The complete list of websites identified following the methodology described above is provided in a table in the supporting information. The information aggregated at the state level is presented in Appendix A, detailing the scores for each state in each of the five categories. A total of 275 websites were found to have data, with a mean number of 5.7 websites per state. The number of sites per state ranges from 0 websites with data in Rhode Island or only one website in West Virginia to a maximum of 15 websites in Texas. Note

Table 2
Number of States With Analytical and Exploration Tools

	Groundwater	Surface water	Quality	Use	Financial
Summary statistics	18	22	10	12	10
Uncertainty	0	0	0	0	0
Projection	2	4	1	4	0
Visualization of data	24	23	24	11	10
Interactive map	38	27	32	14	12
Export capabilities	40	26	32	12	0

Note. Number of states in conterminous United States offering analytical and exploration tools broken down per category of water data.

that depending on the state's objectives, the number of state-level agencies with water management responsibilities and the chosen hosting platform, separate websites do not necessarily indicate that more data are disclosed by the state. It could, instead, reflect a lack of centralization. Out of the total of 275 websites, 213 are from governmental organizations, 51 are from academic institutions, 6 are from the civil society, 2 are from news organizations, 2 are from consortia of academic and governmental organizations, and 1 is from the private sector.

This breakdown is not surprising. State universities (hosting 19% of the websites) are privileged partners of governments for water-related issues and are often mandated by the state to collect and analyze the data, for example, for Delaware, Illinois, Kansas, and Kentucky. One should note, however, that the academic institutions behind the reviewed websites are not necessarily located in the state for which data are presented; for example, the University of North Carolina is the creator and host of water rate dashboards for eight states (and six more have been added since the completion of the survey). Civil society websites are from independent organizations such as Alabama Water Watch, California Data Collaborative, or the New York Public Interest Research Group. The civil society together with consortia are behind many more data websites, but they often follow different boundaries than the state, such as a watershed. The Bear River basin, the Connecticut River, or the North Georgia water associations are examples. In two instances, news organizations also provided state overview of water data; the Texas Tribune provides a statewide interactive map of the reservoir water levels maintained until hurricane Harvey, while StateImpact Oklahoma provided a snapshot of water rates for the state in 2012 also in the format of an interactive map. These illustrate that current water information was deemed relevant to a general audience.

3.1. Review of the Websites' Content

We focus only on whether the information is available for a given state and refer to section 3.2 for a review of the centralization of the various data sets. Table 1 reports the number of states disclosing information in each category and subcategory (the breakdown per state is disclosed in Tables A1 and A2).

Scores for data content are computed for each state and category (i.e., groundwater, surface water, quality, use, and financial) by giving one point for each subcategory of information listed in Table 1 for which one or more websites were identified. Scores for data presentation are computed in a similar fashion for the five categories for each of the indicators listed in Table 2, that is, summary statistics, uncertainty, projection, visualization, interactive map, and export capabilities. The total scores of data content and data presentation are plotted in Figure 1. Key external information on water use, climate, and water quality are reported in Figure 2 to facilitate the interpretation of the results.

3.1.1. Ground and Surface Water Data

Groundwater information is available for nearly all states, with a few exceptions in the northeast (Figures 2a and 2b). Groundwater is arguably of low importance in the northeast because surface water is the primary source for 85% to 100% of water use; however, we note that surface water data are not often encountered for the northeast either. A reason for this might be that the region does not suffer from a lack of water and thus rely on their state-level USGS websites. Demarcation present in the Koppen-Geiger climate classes (Figure 2c) seems to align much more with the groundwater and surface water scores, indicating that water scarcity is potentially a strong incentive to deliver water information.

Yet we note that all of the reviewed states except Rhode Island have their own web-based platform for surface or ground water, and 39 states present both. We remind the reader that USGS state surface water websites were excluded from the survey. This indicates that states intend to host and present this information themselves. We identify several potential reasons that could motivate this:

- Additional data sets are collected at the state level, which a local authority desires to present itself or barriers prevented the aggregation of this data to the federal level.
- State agencies might have designed tools or conducted other postprocessing of the data to provide the most relevant information to their end users, individuals, businesses, and other stakeholders.
- There might be regulatory requirements or legal incentives to disclose the information themselves, such as liability or simply registry purposes.
- There may be political motivations such as showcasing the success of a given policies or preserving an independence from the federal agencies.
- There could be an interest in centralizing the information and breaking down the division between data categories in a way that is not accomplished at the federal level.

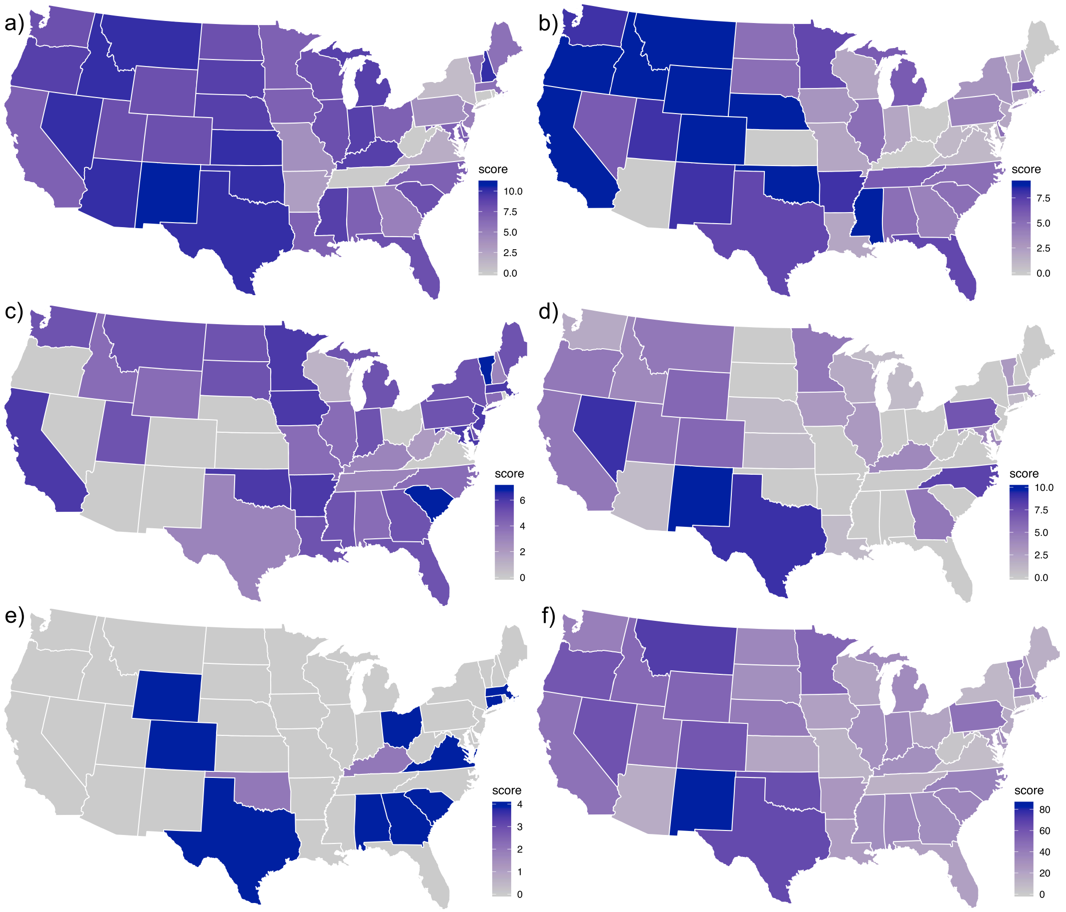

Figure 1. States survey scores for each category: (a) groundwater, (b) surface water, (c) quality, (d) use, (e) finance, and (f) total score. The scores indicate the number of subcategories for which data were found for the given state; the score reflects content, analytics, and exploratory tools but does not inform on the quality and completeness of the hosted data sets.

Though the survey does not allow us to test and evaluate these hypotheses in great detail, the sheer presence of the number of individual state websites suggests that federal platforms present shortcomings.

Further inspection of groundwater platforms shows a great variety of approaches ranging from simple reporting of water tables without any further analysis (see Tables A1 and A2) to sophisticated tools devised to help determine the location and characteristics of new wells by simulating impacts on the water table. Another common type of groundwater-related online resource provides databases of well records or well log information. While key information could be extracted from the logs (e.g., depth of the wells, observed water table, intended use, or permitted withdrawals), it is often reported in PDF forms and thus requires additional processing to render the information usable for analysis. A plausible explanation for this is that these databases are primarily used for administrative purposes.

Surface water data platforms also vary greatly in approach and are quite different from their USGS counterparts. Broadly, websites appear designed to primarily serve one of three purposes: (i) to report discharge measurements with analytics and alternative representations of the data, (ii) to be more transversal and include groundwater information, or (iii) to report permit information. The more informative platforms with the highest scores on the survey aggregate multiple data sets on an interactive map, such as gauge locations, links to historical discharges or water table depth, and permits.

It would be of great value to have platforms that combine the above mentioned approaches to bridge the gap between the infrastructure in place and its use (e.g., wells location and depth, maximum withdrawals, sector of use) and the environmental conditions (e.g., trends in surface and ground water levels). Beyond making the connection between existing data sets and platforms, there is also a need to develop analytics to

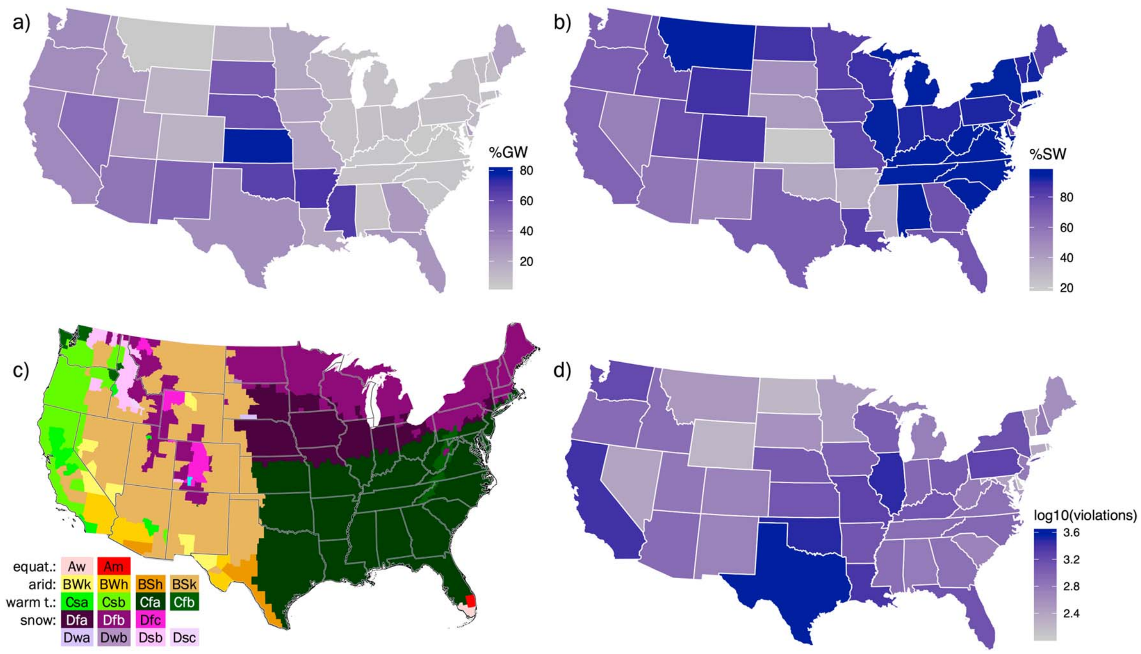

Figure 2. Patterns observed in Figure 1 might show some correlations with the following maps: (a) groundwater irrigation water use per state for the year 2010 (Maupin et al., 2017), (b) percent of total water use supplied by surface water for the year 2010 (Maupin et al., 2017), (c) Koppen-Geiger climate classes per county (Kottek et al., 2017), and (d) water quality violations per state for the period of 1980–2015 (Allaire et al., 2018).

guide interpretation of the measurements, both statistically (e.g., trends, variability, and uncertainty) and phenomenologically (e.g., recharge and climate impacts).

Regarding reservoirs and dams, only 23 states have data platforms that provide some form of information. Around a third of the platforms with reservoir information simply provide time series of water height. The remainder tend to provide a snapshot of the current state of reservoirs, often through maps and representation of the capacity of the reservoirs and their current storage. Rather than simple disclosure of data, these platforms seem to be oriented toward communicating effectively with a broad audience on the current state of storage. This is in stark contrast to state agencies that do not disclose information due to security reasons. Yet nearly half of the surveyed states present storage information. It could be argued that for security reasons it is essential to disclose dam and reservoir characteristics, uses, and maintenance plans so that the public can be made aware of the infrastructural risks. This could enable insurance policies and proper financing of infrastructure maintenance and operations to be developed. Furthermore, the connection with climate is never made explicit, even though it is essential to assess both the relevance of the dams and their services to the ecosystem and population to deal with droughts and floods, and the risk of failures. As for surface and ground water, integrating information across climate, infrastructure, use, and the environment is key to maximize the use and relevance of the platforms.

3.1.2. Water Quality Data

Water quality information is only slightly less common than general surface and ground water data and is presented by 38 of the reviewed states. Only 16 states disclosed return flows permits information, though this information is compulsorily collected by each state with delegated authority for the National Pollutant Discharge Elimination System under the Clean Water Act (45 of the 48 states reviewed). Figure 2d shows the water quality violations reported to the EPA for the period of 1980–2015 aggregated at the state level. There seems to be little correlation between states' reporting of quality measurements and water quality violations reported by drinking water systems.

Some states that have a very high number of violations per person disclose information (with Oklahoma scoring a high mark in the survey), whereas Nebraska has a score of 0, raising the question of access to information pertaining to water standards and subsequently to water justice issues. Underreporting of water quality violations further hinders our capacity to analyze the results and understand states' behaviors.

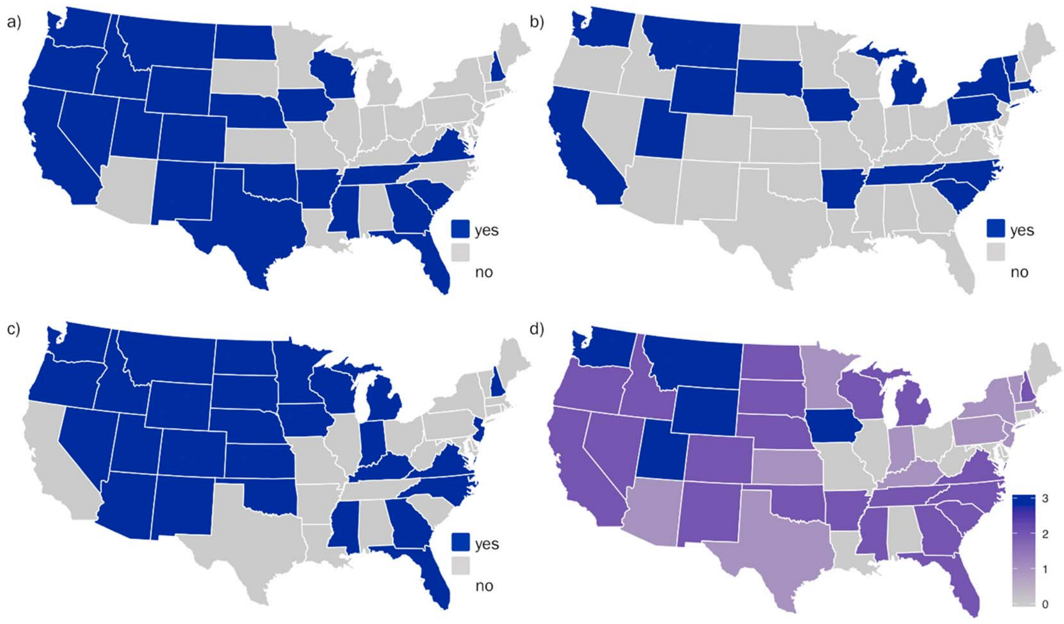

Figure 3. States for which water permits are available for (a) groundwater withdrawals, (b) return flows, (c) surface water withdrawals, and (d) the sum of the three, demonstrating a potential to derive a better understanding of water use.

An examination of the platforms shows three broad types of approach: (i) the disclosure of raw data exclusively, which may be difficult to understand but is transparent; (ii) disclosure of raw data, but centralized on a map to facilitate access to the data; and (iii) more sophisticated visualization of the data with accompanying analytics, in a clear effort to communicate results. It would be interesting to inspect further how each approach was chosen. A more detailed analysis should also be conducted to evaluate the platforms and the data contained within for assessing the successes and failures of the Clean Water Act. It is important to note that no platform makes the connection between quality of raw water used by drinking water systems and quality measurements in groundwater or in-stream locations. An integrated view of water quality issues could provide key insights on infrastructure and financial needs to address the water quality challenges currently faced by the communities, thus bridging objectives of the Clean Water Act and the Safe Drinking Water Act. Furthermore, there is a need to integrate water quality and water quantity considerations. Combining the necessary data sets on a single platform would be a formidable first step to address the water quantity-quality nexus (Gunda et al., 2013).

3.1.3. Water Use

Only 22 states provided some sort of water use information, with as few as five states presenting information as to point of use, raising questions regarding enforcement of water withdrawal permits and allocation. This also challenges the idea that the water footprint of supply chains can be studied systematically without spending tremendous efforts on data collection. Only 14 states provided information as to what sector is using the abstracted water. Figure 1d shows which states disclosed water use information. We observe a pattern that follows the definition of climate zones as illustrated in Figure 2c: Where water is lacking, states made an effort to collect water use information to conduct water budget exercises. The limited information reported at the federal level is thus a consequence of the lack of water use information collected at the state and local levels.

One should note however that though their number is limited, the states that disclose water use information often do so at a much better resolution than available at the national level. For instance, many states provide yearly water use numbers. A flexible data structure at the federal level could allow for more frequent reporting and highlight states that provide yearly numbers. Furthermore, as remarked previously, many states provide databases of water permit information (Figure 3), for groundwater withdrawals (37 states), surface water withdrawals (24), or return flows (16). None of this information is consolidated at the federal level, in particular, no aggregation of the wells location records. It might be possible to leverage this information to derive water withdrawals, use locations, and permitted amounts to inform hydro(geo)logical models.

Some platforms at the state level are already taking one step in this direction, simply by offering interactive maps with points of withdrawals/extractions combined with discharge gauges and observational groundwater wells. Obtaining a similar platform at the national level with a map of both points of use and points of observations would be a fantastic tool for water management and to assess the relevance of the network of observations wells and discharge gauges for the monitoring of water availability.

Despite this untapped source of information, water use data remain scarce in space (rare examples are found at a smaller scale than the county, though arguably water stresses happen at smaller scales; e.g., Mayer et al., 2014) and in time (only a few platforms provide monthly sectoral use). Developing a broader understanding of water withdrawals, consumption and return patterns remains extremely difficult and subject to large uncertainties (Worland et al., 2016). There is also a need to aggregate this information with water availability considerations. Typically this requires streamflow, reservoir storage, and groundwater level data with some processing to be translated into volumes of water available for use. A few platforms can be found that do this, however only at a very local scale (e.g., at the field scale) and without explicit consideration of climate fluctuations. It is essential that data platforms embed climate information to further our understanding and capacity to adapt to climate variability and change.

3.1.4. Financial Information

Water finance data were only found for 11 states. Average water bill information aggregated from survey data is available for some states (e.g., Wyoming). Even when yearly details are publicly accessible, the information is only found within report documents. More sophisticated water rate dashboards, constructed by the University of North Carolina for eight states, permit the comparison of water rates between municipalities as a function of use, rate structure, source of water, and various other facility- or demographic-related characteristics depending on the state for which the platform is built. While the dashboards were initially intended as a benchmarking tool to aid small water utilities, they could inform potential commercial or industrial structures and individual water users as to the water costs they encounter.

Additional water utility financial information of interest, such as operation and maintenance expenses and capital expenditures, were not found, possibly because they are not directly regulated. The only two exceptions are Kentucky and Oklahoma, which present data sets for state-administered infrastructure funding programs with project descriptions and costs related to the water infrastructure financing. It should be noted that the Kentucky Infrastructure Authority's platform is significantly more comprehensive, including budgeting information at the utility level, to support a state initiative to overhaul water infrastructure and encourage consolidation of water systems.

Though a few examples were found, it would be of great interest to expand the water rates dashboards and water infrastructure financing to all states and consolidate the results at the federal level. Indeed, a national understanding of water rates structures and bills at the municipal level, put in perspective with demographic information, systematically performed every year and for every municipality, would not only improve water use information but would support much further analysis of water access issues. Similarly, a national overview of water infrastructure projects and utility maintenance cost would allow researchers and governmental entities to target areas of particular need and to devise much more specific infrastructure financing plans.

3.1.5. Analytical and Exploratory Tools

In terms of tools proposed to analyze the data, around half of the states that provide ground or surface water information provide summary statistics with it. Systematic review of the websites reveals that typical summary statistics are mean, minimum, and maximum values, rarely going into more depth. The presentation of analytics is rare for water quality, though they would be especially relevant in helping users understand trends and significant thresholds for key indicators. This questions our assumption that state-level data are hosted to provide additional analysis. While it might be true for some, it is not systematically the case.

None of the surveyed websites provided uncertainty estimates around the numbers that are presented. This goes against all metrology recommendations. Error estimates would facilitate the aggregation of water data sets by making explicit differences in standards of accuracy, across states and across data categories. Uncertainty propagation is a central element of water security assessments that would further inform which management strategies and policies to adopt in order to mitigate the risks and also simply to evaluate the value of water data and identify collection campaigns that should be conducted.

The number of states proposing projections for any of the categories ranged between 1 (water quality category) and 4 states (surface water and use categories). There might be liability reasons for this, though one should note that projections are regularly proposed for weather and climate information, such as the USDA drought projection. Further inspection of the websites indicates that the time horizon of the projections ranged between 3 days to a century, with a preponderance for long-term projections and assessment of climate change impacts. These projections rarely come hand in hand with uncertainty estimates around the projections. There are only two exceptions: a case where it is possible to compare multiple scenarios at once, thus providing a range of variability, and the second instance where projections for the future are based on extreme values of historical data without attempting forecasts.

As for exploratory tools and the ability to analyze data, the overall impression that we get is that these platforms are not designed to provide a global overview of water within one state but rather for a location-specific use of the data. For example, permit details are often shown one by one. Download capabilities are often restricted to a specific number of entries or a maximum geographic area and rarely permit exports to other tools or online displays (e.g., through an API) for further analyses.

3.2. Centralization of Water Data

The two previous subsections focused on what information is available for each state by aggregating the results at the state level. In this section, we focus on individual websites to evaluate water data centralization potential and to inform federal-level platforms that could constitute a one-stop hub for water data. One hundred sixty-nine websites were found to have data only pertaining to one category, 59 had data for two categories, 40 for three categories, and 3 for four categories. No websites were found to disclose information pertaining to all five categories, not surprisingly due to the lack of financial information. This nonetheless demonstrates that a centralization of water data is possible on a single platform.

The three websites covering four categories of data are the “Ground Water Management Branch Map Interface” from the North Carolina Department of Natural Resources, Utah's “Division of Water Rights Esri Platform,” and Wyoming's “Water and Climate Web Atlas” by the University of Wyoming. Note that the first two are also top global scorers, with scores of 22 and 21, respectively. Three other high-scoring websites are the Colorado Water Conservation Board's “Data Viewer Mapper” (score of 18), the California Department of Water Resources' “Water Data Library” (score of 18), and the Mississippi Water Resources “Data Compendium” (score of 23). These three show data for only three categories, with the last two excluding water use considerations, while Colorado's Data Viewer Mapper leaves aside water quality. They, however, illustrate that sophisticated platforms beyond a simple repository are possible and serve a purpose. Further research should investigate the role they play and who the users are to evaluate their success and learn from them (Laituri & Sternlieb, 2006).

4. Conclusion

Though the survey did not quantify the volume of data available at the state level, it confirmed that the water data gaps identified at the federal level can be filled in part with state-level information. For groundwater, this is currently the object of an initiative led by the USGS, the Advisory Committee on Water Information, which is centralizing information of many universities, states, and other entities to extend the USGS network of water table gauges. The aggregation is still underway as the task presents a tremendous challenge due to differences in standards and accuracy. Indeed, while there are rich and massive data assemblages on water, these data are archived in discrete databases with incompatible units; inconsistent classifications; varied structural, temporal, and spatial organization; and they are maintained by different state and federal agencies (Wiener et al., 2017) or are assembled by water-specific governance bodies (e.g., inter-state compacts), local or county-level government (e.g., water quality for small systems), or consulting groups (e.g., water rates).

The responsibility of centralization does not solely rest on the shoulders of the USGS; many initiatives are underway. Notably, the Food and Agriculture Organization has launched an initiative on water accounting for agriculture to establish the adequate definitions necessary to fill the water gap existing between water supply and irrigation use; and the Aspen institute is leading a dialogue series on a national policy framework to address “the institutional barriers to scaling the integration of water data and information to support sustainable water management” (Patterson et al., 2012) .

However, assuming all aggregation challenges can be addressed, collecting state-level information would not suffice to meet all water data needs. Though it would improve the coverage for groundwater levels, water use data would still remain notably sparse, in particular regarding the identification of point of source, use and returns, making systematic water footprint assessment along supply chains impossible. Beyond federal- and state-level data sets, additional sources of information at different scales could be assembled to improve water use estimates (Dunham et al., 2014). Water reuse is also greatly absent (Wiener et al., 2017). Financial information essential to the construction of hydroeconomics models is equally hard to come by aside from the recent development of the water rate dashboards proposed by the University of North Carolina. This renders the evaluation of the water stress or infrastructure development at the national scale prone to great uncertainties, such that the words of caution voiced by Beider and Tawil (Office2002) regarding the accuracy of their estimates would likely not be very different today.

To fill this data gap, beyond increasing primary data collection, recent developments could be leveraged to derive the quantities of interest using secondary information. A particularly promising lead is remote sensing, such as with GRACE, which provides information that can be used to infer groundwater storage changes, though still at limited spatial resolution (e.g., Doöll et al., 2014, 2018) or, less commonly, water quality (Gholizadeh et al., 2014). Additional remote sensing images and other products at excellent spatial resolution could also be harvested using machine learning tools to identify point of extraction, point of returns, and irrigation systems; quantify water use for outdoor space; or derive irrigation volumes using crop models (Kanwar et al., 2016). Besides the data that can be collected, this also puts in perspective the reluctance of water facilities to disclose water use for privacy reasons (Jerome, 2006), if the information can be inferred using remote sensing (e.g., location of water infrastructure). The Internet of things seems as though it might also be a promising tool to refine water use and needs estimates for the domestic and public sector in particular (Cheifetz et al., 2013; Earth Security Group, 2002; Klein & Oberg, 2010; Patterson et al., 2012; Yang et al., 2010). Regarding industrial water use, disclosure policies, voluntary or not (e.g., 23% of U.S. companies reporting to the Carbon Disclosure Project disclosed water information; CDP, 2015), represent another untapped source of data.

Beyond creating an inventory of existing data at the state level, the survey also revealed the great diversity in what types of information and tools are centralized. Fifteen percent of the reviewed websites showed data for three categories of data or more, illustrating that a centralization of water data within a single structure is possible. As mentioned by Fishman (2016) and Jerome (2006), an example of such a platform already exists at the federal level: The Energy Information Authority, centralizing energy data, offers projections and analytics to inform operators and investors alike. Water maybe presents a greater challenge due to the sheer number of water facilities in the United States in comparison to the number of power plants. However, these could be overcome soon with increasing computational capacities and the development of standards (e.g., Taylor et al., 2017) and automation capabilities. Extensive literature exists on cyberinfrastructures (Castronova et al., 2018; Goodall et al., 2016, 2011; Horsburgh et al., 2011, 2014, Horsburgh & Reeder, 2017; Jeong et al., 2009; Latre et al., 2014; Laituri & Sternlieb, 2006; Larsen & Young, 2013; Laniak et al., 2014; Yang et al., 2018) and more recently the potential of “big data” techniques to assist with this task (Chen & Han, 2017; Klein & Oberg, 2010).

Though we are hopeful that the state of water data in the United States can be improved through the previously mentioned venues, many other entities have a key role to play as the water challenges are very much dependent on other sectors, in particular, the food-energy-water nexus. It is thus necessary to go beyond water data and aggregate demographics-, energy-, food-, and climate-related data (Scanlon et al., 2017). Network information (transportation of energy, water and food, and supply chain characteristics) are also crucial to assess water and infrastructure resiliency. (The university-led project FEWsion is currently undertaking this task for food, energy, and water under NSF Award#1639529). The challenge is only increased by the scale at which data should be reported, both temporally and spatially, ranging from the hour to decades and from the individual to the nation to inform operations and address infrastructure developments. And though time series information is crucial to understand variability and trends, up-to-date concurrent information is equally important to understand current transitions and anticipate future stresses. We further note that confidence intervals and uncertainty estimates around all data measurements are essential for robust propagation of uncertainty and quantification of risks.

Water security has been identified as a key objective by the World Economic Forum. A federal water information platform maintained and managed by a federal water authority could provide a holistic view of water and its governance. This is deemed essential in the context of climate change and higher frequencies of extreme weather, crumbling infrastructure (Beider & Tawil, Office2002), new technologies such as desalinization, transition in the food and energy sector, and growing inequalities in environmental justice. Such federal water authority information could guide policy-makers at various federal agencies to local water utilities, as well as actors from the private sector, such as investors, insurers, manufacturers, and industrials in the food and beverage or mining sector (Earth Security Group, 2016, 2002). It could also inform and support individuals, advocacy groups, civil society, journalists, and educators (Jones & Moulton, 2016). The successes of the decision-makers of tomorrow for the future of the United States depend solely on the quality of the information that will be available to them.

Appendix A: Detailed Scores at the State Level

Tables 1 and 2 present the number of states that disclose each of the categories of data. In this section we disclose the scores per state in Tables A1 and A2.

Table A1
Scores for Each of the Categories and Subcategories Aggregated at the State Level for Alabama to North Carolina

		AL	AR	AZ	CA	CO	CT	DE	FL	GA	IA	ID	IL	IN	KS	KY	LA	MA	MD	ME	MI	MN	MO	MS	MT	NC
	Wells location	3	1	0	3	4	0	4	1	1	1	3	2	4	2	2	1	2	1	1	2	2	1	2	4	2
	Wells depth	3	0	0	1	2	0	4	1	0	1	3	1	3	2	2	1	0	1	1	2	1	0	2	3	2
	Wells water table	3	0	0	1	1	0	4	1	0	0	3	1	3	2	2	1	0	1	1	2	0	1	1	3	2
	Use type	0	1	0	1	4	0	0	1	0	2	1	0	1	2	2	1	0	0	0	1	1	0	1	4	1
	Permit	0	0	0	0	4	0	0	1	0	2	1	0	2	2	2	0	0	0	0	1	1	0	2	1	1
G	Well record	1	0	0	0	1	0	0	1	0	1	2	1	3	2	2	1	0	0	1	1	1	1	1	2	0
W	Analytics	0	0	0	0	0	0	4	0	0	0	2	1	1	1	0	0	1	2	0	1	0	0	0	3	0
	Projections	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Visualization	2	0	0	2	0	0	2	0	0	0	1	2	0	1	2	0	1	2	0	1	0	0	1	4	2
	Interactive map	3	0	0	2	1	0	1	1	0	1	2	2	2	2	2	0	1	2	1	0	1	0	1	4	2
	Download/export	1	1	0	2	2	0	4	1	0	1	3	2	3	2	2	1	1	1	1	1	2	1	2	4	0
	Gauges location	2	3	0	2	6	0	1	1	1	0	2	2	1	0	0	1	2	1	0	1	3	2	2	3	1
	Gauges discharge	1	0	0	1	4	0	1	1	0	0	2	0	0	0	0	1	0	0	0	1	2	0	2	1	1
	Reservoir levels	0	1	0	2	2	1	0	1	0	0	1	1	0	0	0	0	0	0	0	0	2	0	1	2	0
	Use type	0	1	0	1	3	0	0	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	1	2	0
S	SW permit	0	1	0	1	3	0	0	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	1	2	0
W	Analytics	1	1	0	2	1	0	1	0	0	0	2	3	0	0	0	0	1	0	0	1	2	0	2	1	1
	Projections	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Visualization	1	1	0	2	1	0	1	0	0	0	1	3	0	0	0	0	1	0	0	1	2	0	1	2	1
	Interactive map	1	1	0	2	1	0	0	1	0	0	2	2	0	0	0	0	1	0	0	1	2	0	2	2	1
	Download export	0	2	0	3	5	1	1	1	0	0	2	0	0	0	0	0	1	0	0	1	2	1	2	2	0
	Withdrawals	0	0	0	1	2	0	0	0	0	0	1	1	0	0	0	0	0	1	0	0	0	0	0	0	2
	Diversion	0	0	0	1	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
	Point of use	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0
	Water use	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	2
	By sector	0	0	0	0	2	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	1	0	0	2	2
U	Returns	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1
S	Online reporting	0	0	1	1	0	1	0	0	0	0	0	0	0	1	0	1	0	0	0	1	1	0	0	0	0

Table A1 *(continued)*

		AL	AR	AZ	CA	CO	CT	DE	FL	GA	IA	ID	IL	IN	KS	KY	LA	MA	MD	ME	MI	MN	MO	MS	MT	NC
E	Analytics	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	2
	Projections	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
	Visualization	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	1	0	0	0	2
	Interactive map	0	0	0	0	1	0	0	0	0	0	1	0	0	0	1	0	0	1	0	0	0	0	0	1	1
	Download/export	0	0	0	1	2	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	1	0	0	1	0
	Groundwater	0	1	0	3	0	1	1	0	2	0	1	0	2	0	1	0	1	0	0	0	1	1	1	1	3
	Surface water	1	1	0	2	0	1	1	0	2	0	0	1	1	0	0	1	1	2	1	2	5	2	1	2	3
Q	Return permits	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	1
U	Analytics	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	2	1	0	0	0
A	Projections	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
L	Visualization	1	1	0	2	0	1	1	0	1	0	1	1	1	0	0	0	1	1	1	1	4	0	1	0	0
	Interactive map	1	1	0	2	0	0	1	0	1	0	1	1	1	0	1	0	1	2	0	1	6	0	1	2	1
	Download export	1	1	0	3	0	1	1	0	2	0	1	1	2	0	1	1	1	2	1	2	5	1	1	2	0
	Finance	1	0	0	0	1	1	0	0	2	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0
	Analytics	1	0	0	0	1	1	0	0	2	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
	Uncertainty	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
$	Projections	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Visualization	1	0	0	0	1	1	0	0	2	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
	Interactive map	1	0	0	0	1	1	0	0	2	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0
	Download/export	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Note. GW = groundwater; SW = surface water; QUAL = quality; $ = finance.

Table A2
Scores for each of the categories and subcategories aggregated at the state-level for North Dakota to Wyoming

		ND	NE	NH	NJ	NM	NV	NY	OH	OK	OR	PA	SC	SD	TN	TX	UT	VA	VT	WA	WI	WV	WY
	Wells location	4	2	2	1	6	6	0	1	3	3	1	2	3	0	2	2	1	5	4	3	0	3
	Wells depth	1	1	2	0	6	1	1	1	3	2	0	1	3	0	2	2	0	1	2	1	0	3
	Wells water table	1	2	2	0	3	3	0	1	3	2	0	1	2	0	3	2	0	1	1	1	0	2
	GW use type	3	0	2	0	4	4	0	0	2	3	1	1	1	0	2	2	0	1	1	1	0	1
	Permit	2	1	2	1	4	3	0	0	2	1	0	0	1	0	0	2	1	0	2	3	0	1
G	Well record	1	1	2	0	2	2	0	1	2	1	0	1	2	0	2	2	0	1	2	2	0	0
W	Analytics	0	1	1	0	0	1	0	0	1	0	0	1	0	0	2	0	0	0	0	0	0	0
	Projections	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
	Visualization	0	1	1	0	3	3	0	1	2	1	0	0	1	0	3	0	0	0	0	0	0	1
	Interactive map	3	1	1	0	0	2	0	0	3	1	1	1	2	0	3	2	0	1	3	2	0	2
	Download export	3	1	2	1	5	3	0	1	3	3	1	2	2	0	3	1	0	0	2	1	0	2
	Gauges location	3	4	0	0	1	3	0	0	2	4	1	1	2	1	0	2	0	5	3	1	1	3
	Gauges discharge	0	1	0	0	1	1	0	0	1	3	0	0	0	0	0	2	0	0	0	0	0	2
	Reservoir levels	1	3	0	0	2	0	1	0	1	3	0	0	1	1	3	1	0	0	0	0	0	3
	Use type	3	1	1	0	0	2	0	0	1	2	1	0	0	1	0	1	0	0	1	0	0	0
S	SW permit	2	1	1	0	5	1	0	0	1	1	0	0	0	0	1	1	1	0	2	2	0	1
W	Analytics	0	2	0	0	2	0	1	0	1	2	0	0	0	0	3	2	0	0	0	0	0	2
	Projections	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	2
	Visualization	0	1	0	0	2	0	1	0	1	1	0	0	1	0	2	2	0	0	0	0	0	3
	Interactive map	0	2	0	0	2	1	0	0	3	2	1	0	1	0	3	1	0	0	3	0	0	3
	Download/export	3	1	0	0	0	2	0	0	2	3	1	1	1	0	2	0	0	0	1	0	0	1

Table A2 *(continued)*

		ND	NE	NH	NJ	NM	NV	NY	OH	OK	OR	PA	SC	SD	TN	TX	UT	VA	VT	WA	WI	WV	WY
	Withdrawals	0	0	0	0	0	3	0	0	0	0	3	0	0	0	2	0	0	1	0	1	0	0
	Diversion	0	0	0	0	0	2	0	0	0	1	0	0	0	0	1	1	0	0	0	0	0	1
	Point of use	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0
	Water use	0	0	0	0	0	1	0	0	0	1	2	0	0	0	2	1	0	0	0	0	0	2
	By sector	0	0	0	0	0	2	0	0	0	1	4	0	0	0	4	1	0	0	0	0	0	1
U	Returns	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
S	Online reporting	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
E	Analytics	0	0	0	0	0	2	0	0	0	1	5	0	0	0	1	0	0	0	0	0	0	1
	Projections	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	1	0	0	0	0
	Visualization	0	0	0	0	0	1	0	0	0	0	7	0	0	0	1	0	0	0	0	0	0	1
	Interactive map	0	0	0	0	0	2	0	0	0	0	2	0	0	0	0	1	0	1	1	0	0	1
	Download/export	0	0	0	0	3	2	0	0	0	1	7	0	0	0	5	0	0	0	0	1	0	0
	Groundwater	1	0	1	1	0	0	1	0	1	0	0	2	2	0	1	4	0	4	1	1	0	2
	Surface water	1	0	1	1	0	0	1	0	1	0	3	2	2	0	2	2	0	5	1	0	1	2
Q	Return permits	0	0	0	0	0	0	0	0	0	0	1	1	1	2	0	1	0	1	1	0	0	1
U	Analytics	0	0	1	0	0	0	0	0	0	0	1	1	0	0	0	0	0	2	0	0	0	0
A	Projections	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L	Visualization	1	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	2	0	0	0	0
	Interactive map	1	0	0	0	0	0	1	0	1	0	1	1	2	1	1	4	0	1	1	0	0	1
	Download/export	1	0	0	1	0	0	0	0	1	0	2	2	1	1	0	1	0	1	1	0	1	0
	Finance	0	0	0	0	0	0	0	1	2	0	0	1	0	0	1	0	1	0	0	0	0	1
	Analytics	0	0	0	0	0	0	0	1	0	0	0	1	0	0	1	0	1	0	0	0	0	1
	Uncertainty	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
$	Projections	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Visualization	0	0	0	0	0	0	0	1	0	0	0	1	0	0	1	0	1	0	0	0	0	1
	Interactive map	0	0	0	0	0	0	0	1	2	0	0	1	0	0	1	0	1	0	0	0	0	1
	Download/export	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

*Note.*GW = groundwater; SW = surface water; QUAL = quality; $ = finance.

Acknowledgments

The complete list of survey entries is available in the supporting information of this publication. A live version is available on the website (http://awashmodel.org/state-water-data-survey/). We invite the readership to help us maintain, update, and extend the database of water data resources. This work has been accomplished under the NSF award 1360446. Laureline Josset is supported by the Swiss National Science Foundation, grant P2LAP2_161876 and P300P2_171241. Maura Allaire was supported by a NatureNet Science Fellowship, and Chacko Thomas by an Endeavour Postgraduate Scholarship. The authors thank Arthur Goillot and Ain Rosli for their help during the survey process.

References

Allaire, M., Wu, H., & Lall, U. (2018). National trends in drinking water quality violations. *Proceedings of the National Academy of Sciences of the United States of America*, *115*, 207–2083.

Averyt, K., Macknick, J., Rogers, J., Madden, N., Fisher, J., Meldrum, J., & Newmark, R. (2013). Water use for electricity in the United States: An analysis of reported and calculated water use information for 2008. *Environmental Research Letters*, *8*(1), 15001.

Beider, P., & Tawil, N. (2002). Future investment in drinking water and wastewater infrastructure. Congressional Budget Office. ISBN 0-16-051243-3

Blodgett, D., Lucido, J., & Kreft, J. (2015). Progress on water data integration and distribution: A summary of select U.S. Geological Survey data systems. *Journal of Hydroinformatics*, *18*(2), 226–237.

CDP (2018). Closing the gap: Scaling up sustainable supply chains. CDP Supply Chain Report 2017 / 2018, Carbon Disclosure Program.

Castronova, A. M., Goodall, J. L., & Ercan, M. B. (2013). Integrated modeling within a hydrologic information system: An OpenMI based approach. *Environmental Modelling and Software*, *39*, 263–273.

Cheifetz, N., Same, A., Noumir, Z., Sandraz, A. C., Féliers, C., & Heim, V. (2017). Extracting urban water usage habits from smart meter data: A functional clustering approach. In *ESANN 2017, european symposium on artificial neural networks*, pp. 423–428.

Chen, Y., & Han, D. (2016). On big data and hydroinformatics. *Procedia Engineering*, *154*(4), 184–191.

Crutchfield, J. P. (2014). The dreams of theory. *Wiley Interdisciplinary Reviews: Computational Statistics*, *6*(2), 75–79.

De Albuquerque, C. (2011). Report of the Special Rapporteur on the human right to safe drinking water and sanitation. Human Rights Council, United Nations General Assembly.

Dieter, C. A., & Maupin, M. A. (2017). Public supply and domestic water use in the United States, 2015. Open-File Report US Geological Survey, https://doi.org/10.3133/ofr20171131

Dieter, C. A., Maupin, M. A., Caldwell, R. R., Harris, M. A., Ivahnenko, T. I., Lovelace, J. K., et al. (2018). Estimated use of water in the United States in 2015. Circular, US Geological Survey, https://doi.org/10.3133/cir1441

Doöll, P., Müller Schmied, H., Schuh, C., Portmann, F. T., & Eicker, A. (2014). Global-scale assessment of groundwater depletion and related groundwater abstractions: Combining hydrological modeling with information from well observations and GRACE satellites. *Water Resources Research*, *50*, 5698–5720. https://doi.org/10.1002/2014WR015595

Dunham, C., Fuchs, H., & Stratton, H. (2017). Benefits of a national survey on water demand: Existing data and reporting recommendations. Energy Analysis and Environmental Impacts Division, Lawrence Berkeley National Laboratory.

EPA (2002). Data reliability analysis of the EPA Safe Drinking Water Information System/federal version (SDWIS/FED). Report, Office of Water, US Environmental Protection Agency.

Earth Security Group (2016). Earth Security Index 2016: Business Diplomacy for Sustainable Development.

Fishman, C. (2016). Water is broken: Data can fix it. *New York Times*, March 17.

GAO (1990). Drinking water: Compliance problems undermine EPA program as new challenges emerge. Report, Government Accountability Office.

Gallaher, S., & Heikkila, T. (2014). Challenges and opportunities for collecting and sharing data on water governance institutions. *Journal of Contemporary Water Research & Education*, *153*(1), 66–78.

Gholizadeh, M. H., Melesse, A. M., & Reddi, L. (2016). A comprehensive review on water quality parameters estimation using remote sensing techniques. *Sensors*, *16*(8), 1298.

Goodall, J. L., Robinson, B. F., & Castronova, A. M. (2011). Modeling water resource systems using a service-oriented computing paradigm. *Environmental Modelling and Software*, *26*(5), 573–582.

Goodall, J. L., Saint, K. D., Ercan, M. B., Briley, L. J., Murphy, S., You, H., et al. (2013). Coupling climate and hydrological models: Interoperability through Web services. *Environmental Modelling and Software*, *46*, 250–259.

Gunda, T., Hess, D., Hornberger, G. M., & Worland, S. (2019). Water security in practice: The quantity-quality-society nexus. *Water Security*, *6*, 100022.

Harris, M. A., & Diehl, T. H. (2017). A comparison of three federal datasets for thermoelectric water withdrawals in the United States for 2010. *Journal of the American Water Resources Association*, *53*(5), 1062–1080.

High Level Panel on Water (2017). World water data initiative roadmap.

Hirsch, R. M., & Fisher, G. T. (2014). Past, present, and future of water data delivery from the U.S. Geological Survey. *Journal of Contemporary Water Research & Education*, *153*(1), 4–15.

Ho, M., Lall, U., Allaire, M., Devineni, N., Kwon, H. H., Pal, I., et al. (2017). The future role of dams in the United States of America. *Water Resources Research*, *53*, 982–998. https://doi.org/10.1002/2016WR019905

Horsburgh, J. S., & Reeder, S. L. (2014). Data visualization and analysis within a hydrologic information system: Integrating with the R statistical computing environment. *Environmental Modelling and Software*, *52*, 51–61.

Horsburgh, J. S., Tarboton, D. G., Maidment, D. R., & Zaslavsky, I. (2011). Components of an environmental observatory information system. *Computers and Geosciences*, *37*(2), 207–218.

Horsburgh, J. S., Tarboton, D. G., Piasecki, M., Maidment, D. R., Zaslavsky, I., Valentine, D., & Whitenack, T. (2009). An integrated system for publishing environmental observations data. *Environmental Modelling and Software*, *24*(8), 879–888.

Jeong, S., Liang, Y., & Liang, X. (2006). Design of an integrated data retrieval, analysis, and visualization system: Application in the hydrology domain. *Environmental Modelling and Software*, *21*(12), 1722–1740.

Jerome, S. (2016). What feds must do to get a handle on water data, Federal Computer Week, July 28.

Jones, P. A., & Moulton, A. (2016). The invisible crisis: Water unaffordability in the United States. *Unitarian Universalist Service Committee*.

Kanwar, R., Narayan, U., & Lakshmi, V. (2010). Web service based hydrologic data distribution system. *Computers and Geosciences*, *36*(7), 819–826.

Klein, D. R., & Oberg, G. (2017). Using existing municipal water data to support conservation efforts. *Journal of American Water Works Association*, *7*(109), 313–319.

Kottek, M., Grieser, J., Beck, C., Rudolf, B., & Rubel, F. (2006). World map of the Köppen-Geiger climate classification updated. *Meteorologische Zeitschrift*, *15*(3), 259–263.

Laituri, M., & Sternlieb, F. (2014). Water data systems: Science, practice, and policy. *Journal of Contemporary Water Research & Education*, *153*(153), 1–3.

Laniak, G. F., Olchin, G., Goodall, J., Voinov, A., Hill, M., Glynn, P., et al. (2013). Integrated environmental modeling: A vision and roadmap for the future. *Environmental Modelling and Software*, *39*, 3–23.

Larsen, S. G., & Young, D. (2014). WaDE: An interoperable data exchange network for sharing water planning and use data. *Journal of Contemporary Water Research & Education*, *153*(1), 33–41.

Latre, M., Lopez-Pellicer, F. J., Nogueras-Iso, J., Béjar, R., Zarazaga-Soria, F. J., & Muro-Medrano, P. R. (2013). Spatial data infrastructures for environmental e-government services: The case of water abstractions authorisations. *Environmental Modelling and Software*, *48*, 81–92.

Mack, E. A., & Wrase, S. (2017). A burgeoning crisis? A nationwide assessment of the geography of water affordability in the United States. *PloS one*, *12*(1).

Maupin, M. A., Kenny, J. F., Hutson, S. S., Lovelace, J. K., Barber, N. L., & Linsey, K. S. (2014). Estimated use of water in the United States in 2010 Circular 1405. Circular, US Geological Survey.

Mayer, A., Mubako, S., & Ruddell, B. L. (2016). Developing the greatest blue economy: Water productivity, fresh water depletion, and virtual water trade in the great lakes basin. *Earth's Future*, *4*(6), 282–297.

National Academy of Sciences (2002). Estimating water use in the United States: A new paradigm for the national water-use information program. National Academies Press.

National Research Council (2012). Preparing for the third decade (Cycle 3) of the National Water-Quality Assessment (NAWQA) Program, Cycle 3.

National Research Council (2012). *Water reuse: Potential for expanding the nation's water supply through reuse of municipal wastewater*: National Academies Press.

Patterson, L., Doyle, M., King, K., & Monsma, D. (2017). Internet of water: Sharing and integrating water data for sustainability. Report from the Aspen Institute Dialogue Series on Water Data, The Aspen Institute.

Perrone, D., Hornberger, G., van Vliet, O., & van der Velde, M. (2015). A review of the United States' past and projected water use. *Journal of the American Water Resources Association*, *51*(5), 1183–1191.

Ruddell, B. L. (2018). Hess opinions: How should a future water census address consumptive use? (And where can we substitute withdrawal data while we wait?) *Hydrology and Earth System Sciences*, *22*(10), 5551–5558.

Russo, T. A., & Lall, U. (2017). Depletion and response of deep groundwater to climate-induced pumping variability. *Nature Geoscience*, *10*(2), 105–108.

Scanlon, B. R., Ruddell, B. L., Reed, P. M., Hook, R. I., Zheng, C., Tidwell, V. C., & Siebert, S. (2017). The food-energy-water nexus: Transforming science for society. *Water Resources Research*, *53*, 3550–3556. https://doi.org/10.1002/2017WR020889

Shaffer, K. (2009). *Variations in withdrawal, return flow, and consumptive use of water in Ohio and Indiana, with selected data from Wisconsin, 1999–2004*. US Geological Survey.

Sprague, L. A., Oelsner, G. P., & Argue, D. M. (2017). Challenges with secondary use of multi-source water-quality data in the United States. *Water Research*, *110*, 252–261.

Taylor, P., Cox, S., Walker, G., Valentine, D., & Sheahan, P. (2014). WaterML2.0: Development of an open standard for hydrological time-series data exchange. *Journal of Hydroinformatics*, *16*(2), 425. https://doi.org/10.2166/hydro.2013.174

Vörösmarty, C., Askew, A., Grabs, W., Barry, R. G., Birkett, C., Döll, P., et al. (2001). Global water data: A newly endangered species. *Eos*, *82*(5), 54–54.

Walton, B. (2017). UN expert connects U.S. water and sanitation struggles to poverty. Circle of Blue, December 15.

Wiener, M. J., Jafvert, C. T., & Nies, L. F. (2016). The assessment of water use and reuse through reported data: A US case study. *Science of the Total Environment*, *539*, 70–77.

Worland, S. C., Steinschneider, S., & Hornberger, G. M. (2018). Drivers of variability in public-supply water use across the contiguous United States. *Water Resources Research*, *54*, 1868–1889. https://doi.org/10.1002/2017WR021268

Yang, C., Raskin, R., Goodchild, M., & Gahegan, M. (2010). Geospatial cyberinfrastructure: Past, present and future. *Computers, Environment and Urban Systems*, *34*(4), 264–277.

Yang, L., Yang, S. H., Magiera, E., Froelich, W., Jach, T., & Laspidou, C. (2017). Domestic water consumption monitoring and behaviour intervention by employing the internet of things technologies. *Procedia Computer Science*, *111*, 367–375.